Spring 5-2005

Investigation of State lotteries and Poverty

Kevin William Olberding
University of Tennessee - Knoxville

UNIVERSITY HONORS PROGRAM

SENIOR PROJECT - APPROVAL

Name: Kevin Olberding

College: Arts and Sciences Department: Statistics

Faculty Mentor: Dr. William Seaver

PROJECT TITLE: Investigation of State lotteries and Poverty

I have reviewed this completed senior honors thesis with this student and certify that it is a project commensurate with honors level undergraduate research in this field.

Signed: W. L. Seaver, Faculty Mentor

Date: 5-07-05

Comments (Optional):

Kevin Olberding

Investigation of State Lotteries and Poverty

I am investigating the relationship between per capita state lottery sales and poverty. I am also checking to see if median income, education level, or unemployment level for the state has any factor on lottery sales. Previous research, including *Hitting the Jackpot or Hitting the Skids: Entertainment, Poverty, and the Demand for State Lotteries* by Blaylock, Just and Simon (December 14, 2004), have found a positive relationship between state lottery sales and the state's poverty level. This would imply that states with a poorer population spend more money per person playing the lottery.

There has been much speculation that lotteries are a tax on the poor, who play the lottery because "low-income consumers may view lotteries as a convenient and otherwise rare opportunity for radically improving their standard of living. Bad times may cause desperation and the desperate may turn to lotteries in an effort to escape hardship" (Blaylock 1). I will attempt to validate or reject this claim by using statistical methods.

The dependent variable will be Sales, which represents the total sales from the state lottery in dollar amounts divided by the states population. The independent variables will be education, income, poverty and unemployment. Education is the percentage of persons 25 years old and over with a bachelor's degree or more in the state. It would be expected that the higher the education level of a state, the less people from the state would play the lottery because they may have a better understanding of probability theory. Income is the median household income for the state. Income should increase lottery sales, being a normal good, because the more money the population has, the more money it can spend on lottery tickets. Poverty is the percentage of people below the poverty level for the state and as claimed in paragraph 2 should show a positive relationship to sales. Finally, unemployment is the percentage of people unemployed in the state and may have a positive relationship with sales based on the same logic as the poverty variable.

The model was to be as follows:

$$\textbf{Sales} = B_1\textbf{Education} + B_2\textbf{Unemployment} + B_3\textbf{Income} + B_4\textbf{Poverty} + \boldsymbol{e}$$

Due to a co-linearity problem between the variables, a variable had to be removed. To determine which variable to remove, a variable selection routine report was run and used to construct the following model:

$$\textbf{Sales} = B_1\textbf{Education} + B_2\textbf{Income} + B_3\textbf{Poverty} + \boldsymbol{e}$$

This model also had a problem in that there were many outliers in the data. In particular, Delaware, Massachusetts, and Rhode Island have unusually large sales per capita compared to other states with a lottery. To correct the model for this outlier issue, robust multiple regression using Tukey's Biweights was used. This technique down weights outliers in the data set in order to achieve more accurate variable coefficients.

The model did prove to be significant with a p value of .0001. The R^2 was.1326, indicating that 13.26% of the variation in sales per capita is explained by the independent variables education, income, and poverty. In addition, the coefficient for income was significant with a p value of .0006. The coefficient on income was .0063, indicating that if a states median household income increases by one thousand dollars, then on average holding other variables constant, an increase of 6.3 dollars per person will be seen in state lottery sales. This was to be expected, as it has been shown that the lottery is a normal good. The coefficients for poverty and education did not prove to be significant with p values of .2121 and .2134, respectively.

The regression assumptions for residuals were also checked. The residuals were not normally distributed due to the many outliers in the data set. The residuals did, however, show white noise, which was checked using the Portmanteau Test. In addition, the ACF and PACF plots of the residuals were fairly clean.

Since the residuals from the robust model are not normally distributed the results had to be validated using a bootstrap method. By bootstrapping confidence intervals for the coefficients of the variables, it was found that the results of the robust model are valid, and that income's coefficient is positive and significantly different from zero.

The data was obtained from the U.S. Census Bureau website. The data is a combination of years 2000, 2001, 2002, and 2003 for the 37 states that had a lottery during those years giving a total sample size of 148.

It was also found that the data used by *Hitting the Jackpot or Hitting the Skids: Entertainment, Poverty, and the Demand for State Lotteries* had some irregularities in it. For the most part, their data was consistent with the data used in this study, but for per capita sales of Rhode Island's state lottery, their data significantly differed from the data used in this study. This is particularly interesting because Rhode Island has unusually large per capita sales (approx. 1000) but a relatively low poverty level, but this was not expressed in *Hitting the Jackpot or Hitting the Skids: Entertainment, Poverty, and the Demand for State Lotteries*'s data set. Instead, they have Rhode Island's per capita sales as only $193.51, which appears to be extremely underestimated.

Further research, including looking at state lottery sales based on individual counties within the state, could produce more statistically significant results.

In reality, this is a longitudinal study that should have been analyzed as a repeated measures data set, but that topic is beyond the scope of undergraduate statistics.

References:

Blaylock, Just and Simon. 2004. *Hitting the Jackpot or Hitting the Skids: Entertainment, Poverty, and the Demand for State Lotteries*

U.S. Census Bureau:
Income and Apportionment of State-Administered Lottery Funds: 2000
<http://www.census.gov/govs/state/00lottery.html> Accessed 2/26/05
Income and Apportionment of State-Administered Lottery Funds: 2001
<http://www.census.gov/govs/state/01lottery.html> Accessed 2/26/05
Income and Apportionment of State-Administered Lottery Funds: 2002
<http://www.census.gov/govs/state/02lottery.html> Accessed 2/26/05
Income and Apportionment of State-Administered Lottery Funds: 2003
<http://www.census.gov/govs/state/03lottery.html> Accessed 2/26/05
Uncle Sam's Reference Shelf
<http://www.census.gov/statab/www/ranks.html> Accessed 2/26/05

Bootstrap Section

Estimation Results		Bootstrap Confidence Limits		
Parameter	**Estimate**	**Conf. Level**	**Lower**	**Upper**
Intercept				
Original Value	-335.6818	0.9500	-575.9205	-55.4569
Bootstrap Mean	-338.4282			
Bias (BM - OV)	-2.7464			
Bias Corrected	-332.9354			
Standard Error	133.0824			
B(educ)				
Original Value	1.1561	0.9500	-8.0916	11.0843
Bootstrap Mean	0.9603			
Bias (BM - OV)	-0.1959			
Bias Corrected	1.3520			
Standard Error	4.9371			
B(income)				
Original Value	0.0091	0.9500	0.0021	0.0158
Bootstrap Mean	0.0093			
Bias (BM - OV)	0.0002			
Bias Corrected	0.0089			
Standard Error	0.0035			
B(poverty)				
Original Value	7.5936	0.9500	-1.3905	15.1947
Bootstrap Mean	7.7403			
Bias (BM - OV)	0.1466			
Bias Corrected	7.4470			
Standard Error	4.1885			

Kevin Olberding

Description of Independent Variables

Variable Name (expected sign)	Description	Sample Mean (standard deviation)
Poverty (-)	Poverty is the percentage of people below the poverty level for the state	10.768 (2.809)
Income (+)	Income is the median household income for the state	43501.05 (5929.504)
Education (-)	Education is the percentage of persons 25 years old and over with a bachelor's degree or more in the state	27.454 (4.896)
Unemployment (-)	Unemployment is the percentage of people unemployed in the state	5.576 (1.025)

Kevin Olberding

Estimated Sales Model

Dependent variable is the sales per capita of state lottery tickets	
Variable	Coefficient
Constant	-125.2046 (.1232)
Poverty	3.3410 (.2121)
Income	0.0063* (.0006)
Education	-2.1931 (.2134)
# observations	148
R^2	0.1326
F - statistic	7.339* (.0001)
p values are in parenthesis * denotes statistical significance at the 1% level	